University of Nevada, Reno

Geometric Approach to Segmentation and Protein Localization in Cell Cultured Assays

A thesis submitted in partial fulfillment of the

requirements for the degree of Master of Science in

Computer Science

By

Sreevatsan Raman

Dr. Bahram Parvin and

Dr. George Bebis

Thesis Advisors

December, 2005

Abstract

Cell-based fluorescence imaging assays are heterogeneous requiring collection of a large number of images for detailed quantitative analysis. Complexities arise as a result of variation in spatial non uniformity, shape, overlapping compartments, and scale. A new technique and methodology has been developed and tested for delineating sub-cellular morphology and partitioning overlapping compartments at multiple scales. This system is packaged as an integrated software platform for quantifying images that are obtained through fluorescence microscopy. Proposed methods are model-based, leveraging inherent geometric shape properties of sub-cellular compartments and corresponding protein localization. From the morphological perspective, convexity constraint is imposed to delineate, partition and group nuclear compartments. From the protein localization perspective, radial symmetry is imposed to localize punctuate protein events at sub-micron resolution. This technique has been tested against two classes of punctuate protein events centrosomes and foci. This technique has been tested against images that were generated to study centrosome abnormalities in breast cancer cells and DNA double strand breaks in human epithelial cells.

Acknowledgements

There have been many people that in one way or another, knowingly or not, have helped me reach my goal of obtaining a Master's in Computer Science. I would like to first thank my advisor, Dr. George Bebis, for guiding me during all these years. I would like to express my sincere thanks and profound gratitude to Dr. Bahram Parvin, Lawrence Berkeley National lab, for his ideas and inspiration which helped me move when I got stuck. I would like to thank Dr. Christopher Alan Maxwell, Lawrence Berkeley National lab, for his support and help in testing out the software and his comparative analysis. I would also like to thanks Dr. Mary-Helen Barcellos, Lawrence Berkeley National lab, for her support. Thanks to Dr. Mircea Nicolescu and Dr. Mark Pinsky for taking part in my committee. Thanks to my mother Vasumathy Raman and my friends for their support and motivation that helped me reach where I am.

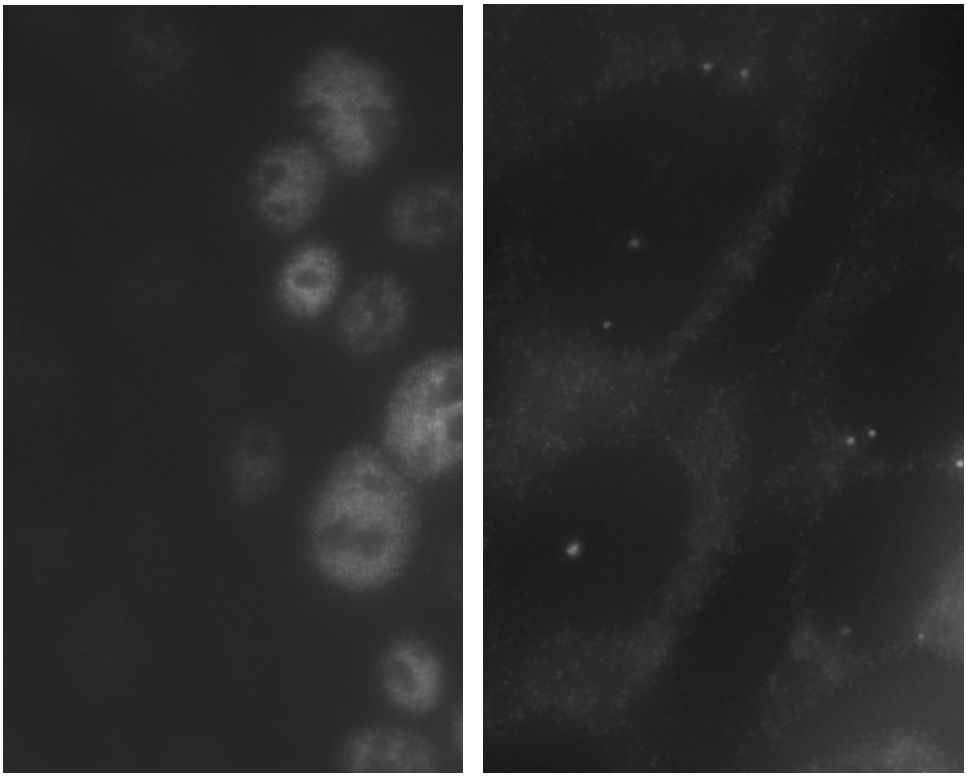
This work was supported by Low Dose Radiation Research Program, Biological and Environmental Research (BER), U.S Department of Energy.

Table of Contents

List of figures

Chapter 1: Introduction

Response of tissues and biological material to various stimuli usually requires collecting and analyzing large samples of data for each experimental variable e.g., tissue type, dosage, type of stimuli and concentration. The response can be multi spectral and multi dimensional and can be imaged used different microscopic techniques. Quantitative analysis of these responses is a necessary step towards visualization of large co-localization studies. The sub-cellular location of the protein expression is an important property that provides the context in which the protein carries out its function. Protein expression may be diffuse or punctuate as shown in figure 1. When the protein expression is diffuse the average value of the signal within a particular context is measured. When the protein expression is punctuate additional step is needed within the specific context for quantitative assessment. This work encompasses a complete methodology and quantitative assessment of co-localization studies in cell cultured assays. This work has been developed for studying quantitative assessment of protein localization in a wider spectrum of co-localization studies that has a variety of image signature, scale and complexities. The complexities are addressed using approaches that leverage the inherent geometry of the quantity that is measured.



(a) (b)

Fig 1: Protein expression: (a) Diffuse (b) Punctate

Fluorescence microscopy is a widely used imaging technique for cell cultured assays. Analyzing large sets of data for detailed quantitative analysis visually is a cumbersome process and therefore there is a need to automate the process. Automating the analysis offers another advantage that the assessment is objective than subjective. This technique has been tested against studying centrosomal abnormalities (CA) and analyzing foci associated with DNA double strand breaks and it can be extended to other phenotypic studies.

Centrosomes are major microtubular organizing center in an animal cell. It comprises of two orthogonally arranged centrioles surrounded by peri centriolar material. They play a vital role during cell division. During DNA synthesis phase of a cell cycle centrosome replicates itself. The replicated centrosomes separate and nucleate a bipolar spindle that equally contracts and segregates the replicated information into daughter cells during the mitosis phase.

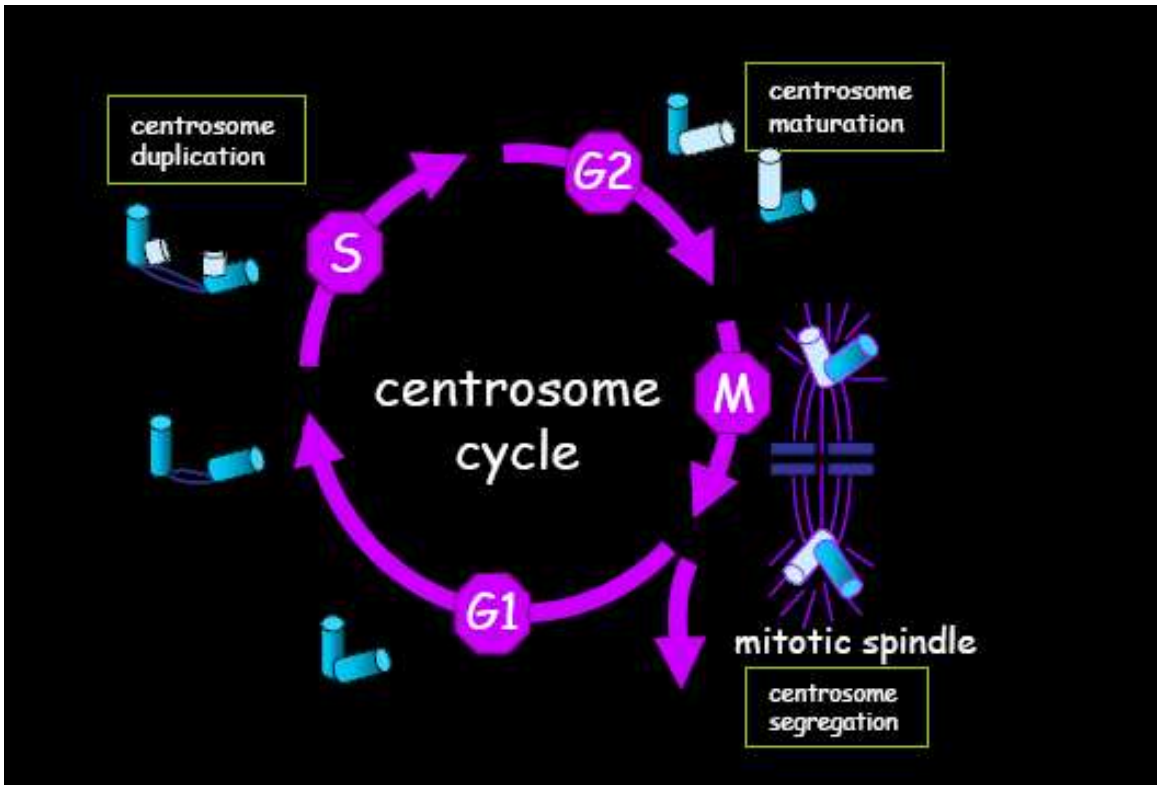


Fig 2. Cell Cycle

One facet of centrosomal abnormality (CA) refers to the presence of more than two centrosomes in a cell which leads to abnormal cell division. As CA is a rare event a large number of data needs to be collected for each experimental variable for an objective

result. Complexities arise due to non uniform staining, overlapping nuclear regions, touching centrioles and scale of the nuclear regions.

In the proposed system complexities are addressed using a model based approach which uses inherent geometric features of the sub-cellular nuclear compartments and punctuate protein events for quantitative assessment. The geometric features take into account the convexity features of the nuclear compartments and radial symmetry of the protein events. Nuclear segmentation is performed by higher level geometric analysis of edge fragments which are obtained by differential operators. In cell cultured assays imaged for fluorescence microscopy some nuclei are isolated and some nuclei are clumped. The strategy is to delineate isolated nuclei first and then to decompose clumped nuclei subjected based on iterative decomposition. The given non-convex blob is partitioned into optimum convex blobs by iterative decomposition based on constraint stratifying network. The constraints ensure that the computational cost is minimum and the optimum partition of the configuration is obtained which represents the sub-cellular compartments more accurately.

Protein localization is performed by iterative voting[13], which are kernel based and the topography favors radial symmetry. The method implemented here falls into the category of iterative techniques which are adaptive to geometric perturbation and typically produce more stable results. This method shares several attributes with tensor-based voting, but it differs in that it is scalar and iterative. Voting along the gradient direction provides a hypothesis of saliency which is very coarse, updating and reorienting the kernel every iteration enables and edge location spatial saliency can be inferred. The kernel provides an approximate centre of mass that is coarse initially. However, it is

updated and refined at every iteration to collapse the object to its center of mass which is picked up by a voting threshold. It is robust with respect to variations in scale and intensity.

Chapter 2: Previous Work

The difficulties in localization of subcellular compartments are often due to variations in scale, noise, and topology. Other complexities originate from missing data and perceptual boundaries that lead to diffusion and dispersion of the spatial grouping in the object space. Techniques for extraction of nuclear compartments vary from, global thresholding or adaptive (localized) thresholding followed by watershed method [11] for separating adjacent regions, curve evolution or shape regularization of sub-cellular compartments [4,5,12] to geometric techniques using nonlinear diffusion[1,12] . Techniques in radial symmetries, as evident by protein configuration, can be classified into three different categories: (1) point operations leading to dense output, (2) clustering based on parameterized shape models or voting schemes, and (3) iterative techniques.

Point operations are usually a series of cascade filters that are tuned for radial symmetries. These techniques use image gradient magnitudes and orientations to infer the center of mass for regions of interest [6,7,10]. Parametric clustering techniques are often based on a variant of the Hough transform, circle or ellipse finders. These techniques produce loci of points corresponding to the parametric models of well-known geometries. These point distributions are then emerged, and model parameters are refined. Non-parametric clustering techniques operate along the gradient direction to search for radial symmetry, using either line- or area-based search. Line-based search is also known as the spoke filter, where the frequency of occurrence of points normal to the edge direction is aggregated. In contrast, area-based voting accumulates votes in a small

neighborhood along the gradient direction. Examples of iterative methods include the watershed method and the regularized centroid transform (RCT)[12], which transport boundary points to the local center of mass iteratively. These can be classified as curve-based voting since the voting path is not along a straight line but along a minimum energy path. Voting paths can be easily distorted by noise, local structures, and other singularities in the image, and may lead to over-segmentation. Thus, the problem is often regularized at different levels through either non-linear diffusion of random noise, non-linear diffusion of speckle noise, or enforcing smoothness of the path leading each point on the surface to its local centroid. Parametric techniques tend to be more robust as long as the geometric model captures pertinent shape features at a specific scale, e.g., Hough transform. Iterative methods, such as watershed, regularized centroid transform, and geometric voting, produce superior results because they compensate for larger variation of shape features.

The first two categories of radial symmetry detection can be summarized as follows. Interest-point operators are fast and well-suited for detecting small features for higher levels of interpretation and manipulation. Parametric voting techniques are potentially memory intensive, depending upon the dimensionality of the parameter space, and remain sensitive to small deviations from the underlying geometric model. Line- and area-based voting produces a voting space that is diffuse and subject to further analysis.

Chapter 3: Segmentation

Cellular compartments are stained to be identified in a 2D cell cultured assays imaged for fluorescence microscopy. The flourophore for nuclear stain in a fluorescence microscopy responds to a particular wavelength of light and thus illuminating the sample for a particular wavelength. This illumination is imaged as the nuclear compartments for subsequent analysis.

The nuclear compartments that are imaged are sometimes isolated and are sometimes clumped. The strategy for delineating nuclear regions is to extract the isolated nuclear compartments first and to apply additional geometric constraints for clumped nuclear regions. Though the image signature might suggest that a simple thresholding based approach could delineate the sub cellular compartments, the variations in shapes, non uniform staining and imaging artifacts demands a more robust and localized strategy. The localized strategy is an edge based approach with a constraint satisfaction network based on geometry of the compartments for delineation. The edges are obtained from a Laplacian operator; the edge information is combined with the gradient information to extract closed contours. These closed contours are subjected to convexity tests. If the convexity test fails then the contour is subjected to further steps for delineation through a constraint satisfying network. The steps are given as follows:

3.1 Boundary extraction

The boundary extraction is initiated from zero-crossings of the Laplacian operator. Let I(x,y) be a 2D image; the zero-crossing edge information is extracted by using a Laplacian operator which is a second derivative operator.

This edge information is coupled with gradient information at the same scale; two thresholds of gradient high threshold and low threshold are used to filter the edge information. The zero crossing ensures that boundary is closed and continuous as compared to standard approach by canny edge detection. The threshold high and low ensures that spurious contours are eliminated. The edge information is linked from high to low thresholds to form a closed boundary.

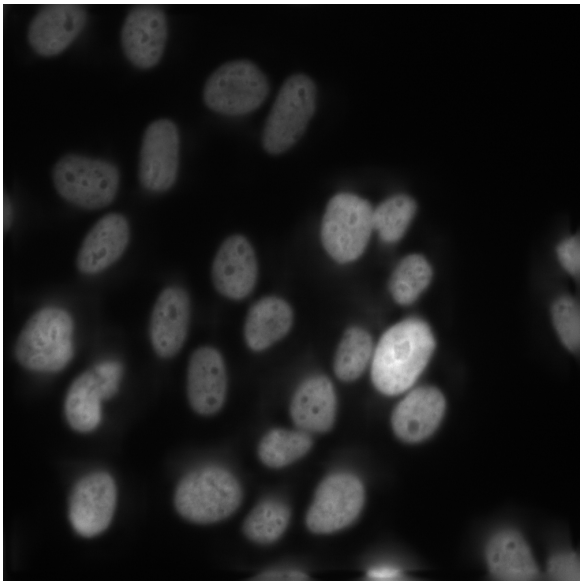


Fig 3. Original image

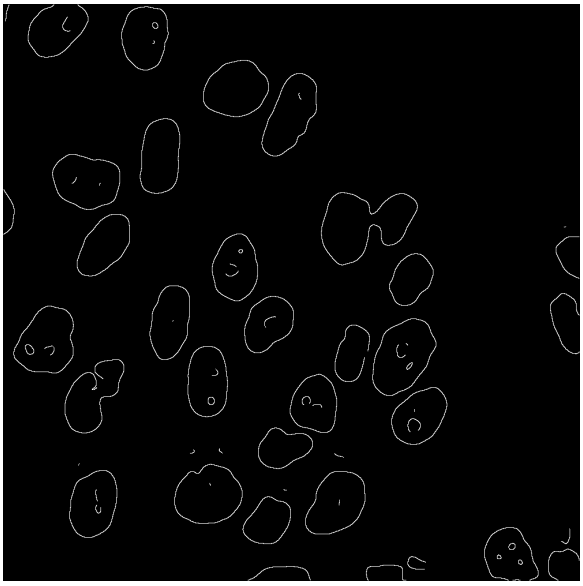


Fig 4. Zero crossings High Threshold gives discontinuous contours

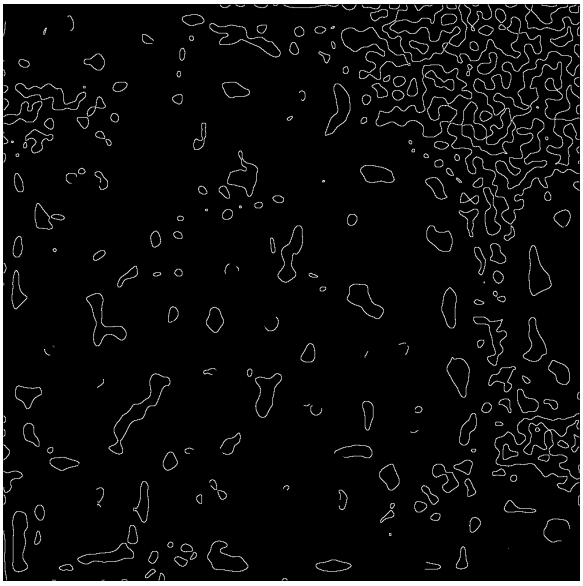


Fig 5. Zero Crossings Low Threshold

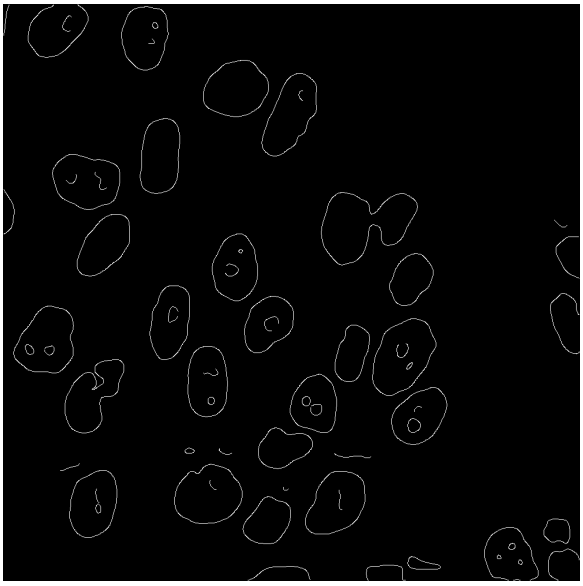


Fig 6. Zero crossing boundary extraction from high threshold to low threshold

3.2 Convexity Test

The contours extracted from the pervious step is subjected to convexity test to determine if the nuclear compartment is isolated or clumped which needs additional processing for delineation. The convexity test is performed by approximating the contour as a polygon and measuring the total turning angle around the polygon to determine the convexity. If the nuclear region is isolated the total angle measured for the approximated polygon is 360 degrees; for a clumped nuclear region the measured angle greater than 360 degrees which indicated that additional steps are needed to delineate the contour which represents a clumped nuclei.

3.3 Grouping and Partitioning

3.3.1 Grouping

The contours which fail the convexity test as explained in section 4.2 is subjected to partitioning to obtained individual nuclear compartments. The partitioning algorithm is based on iterative decomposition and constraint satisfaction. The partitions are initiated from points of positive curvature maxima. The iterative decomposition ensures optimum number of nuclear compartments is obtained and constraint satisfaction ensures that computation cost is the least. The following constraints are used to minimize the computational cost

A) Positive Curvature Constraint

The positive curvature constraint ensures that the partition is initiated from the folds of the contours which are the ideal points for potential decomposition. The curvature of the contour is calculated by the formula

$$k = \frac{\delta' x \delta'' y - \delta' y \delta'' x}{(\delta' x^2 + \delta' y^2)^{3/2}} \quad (1)$$

The contour derivative is computed by convolving the contours information with derivative of Gaussian function.

(B) Anti Parallel Constraint

Anti parallel constraint ensures that the each pair of positive curvature point i.e., points of potential decomposition is anti parallel. The anti parallel constraint is asserted by computing tangent direction for each point in a small neighborhood. The anti parallel

constraint reduces the computational cost of the algorithm as it reduces the number of hypothesis for a possible partition.

(C) Non Intersecting Constraint

The non intersecting condition asserts that the partitions do not cross the boundary of the contour or hypothesized partitions.

(D) Convexity Constraint

Nuclear compartments are always convex and the convexity constraint ensures that partitions obtained from the iterative decomposition are convex. This constraint avoids incorrect segmentation.

3.3.2 Partitioning

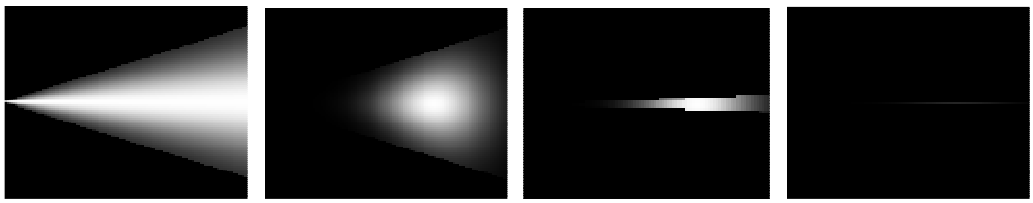
Each set of clumped nuclei is partitioned by using the constraints mentioned in sections 3.3.1 (A) to (D). Each configuration has a cost function which is given by

$$\sum_{i=1}^{n} \frac{\phi_i - 2\Pi}{2\Pi} \qquad (2)$$

The optimum configuration is given by the least cost function as in equation 2 satisfying all the constraints. The decomposition algorithm is summarized in appendix A.

Chapter 4: Protein Localization

The segmentation of the nuclear regions gives the context for measuring the punctuate protein events. Protein events in a cell cultured assay may be diffused or punctuate. If it is diffused then the value is averaged. If the protein events are punctuate then the protein events are isolated using an iterative voting technique. The strategy to localize the protein events is to locate the centre of mass of the protein events through iterative voting technique which favors radial symmetry. Voting long the gradient direction gives the hypothesis of saliency which is initially coarse. However, refining the kernel at each iteration spatial saliency can be inferred. The kernel provides a prior knowledge for an approximate location of center of mass, which is subsequently refined as the shape of the kernel becomes more focused, as shown in Figure 7.



(a) (b) (c) (d)

Fig 7. Kernel Topography: (a)-(d) Kernel evolution to detect radial symmetries

Let I(x, y) be a 2D image. Let $\alpha(x, y)$ be the voting direction at each image point, where

$$\alpha(x, y) = (\cos(\theta(x,y)), \sin(\theta(x,y)))$$

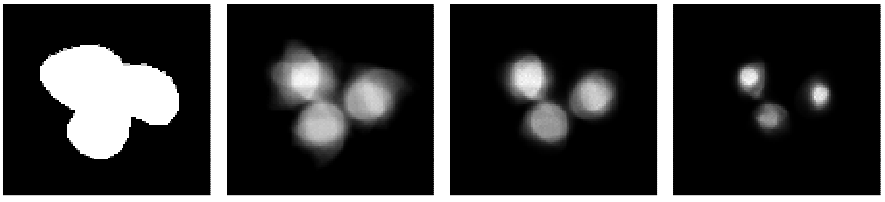
for some angle $\theta(x,y)$ that varies with the image location. Let $\{r_{min}, r_{max}\}$ be the radial range and Δ be the angular range. Let $V(x,y; r_{min}, r_{max}, \Delta)$ be the vote image, dependent on the radial and angular ranges and having the same dimensions as the

original image. Let A(x,y; r_{min}, r_{max}, Δ) be the local voting area, defined at each image point (x,y) and dependent on the radial and angular ranges, defined by

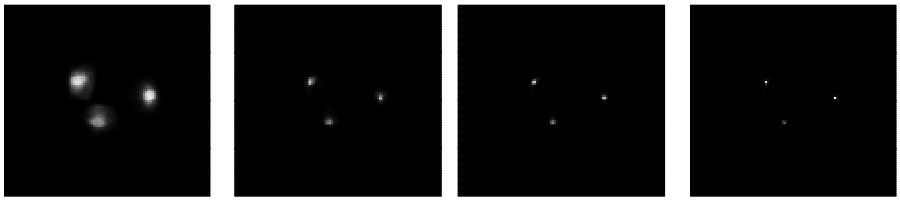
$$A(x, y; r_{\min}, r_{\max}, \Delta) := \{(x \pm r\cos\phi, y \pm r\sin\phi) \mid r_{\min} \leq r \leq r_{\max} \text{ and } \theta(x, y) - \Delta \leq \phi \leq \theta(x, y) + \Delta\} \quad (3)$$

Finally, let K(x, y; σ, α, A) be a 2D Gaussian kernel with variance σ, masked by the local voting A(x,y; r_{min}, r_{max}, Δ) and oriented in the voting direction α(x,y). Figure 7, shows a subset of voting kernels that vary in topography, scale, and orientation. The voting algorithm projects the gradient information along the edges with an initial large spread function. At each step of iteration, the total kernel energy is more focused along a specific orientation corresponding to the local maximum of the voting landscape. This technique collapses edge information radially to a single point, which is close to center of mass of each distinct object.

An example of the application of radial kernels to overlapping objects is shown in Figure 8 together with the intermediate results. The voting landscape corresponds to the spatial clustering that is initially diffuse and subsequently refined and focused into distinct islands. The iterative voting scheme is summarized in Appendix B



(a) (b) (c) (d)



(e) (f) (g) (h)

Fig 8. Voting: (a) Synthetic image simulating three overlapping protein events (b) – (h) voting landscape at each iteration

Chapter 5: BioQuant

The algorithms presented for extracting sub-cellular compartments and protein localization in chapters 3 and 4 has been packaged into a software package (BioQuant) which is used to quantify protein events in cell cultured assays. BioQuant is built to automatically quantify information and produce the result of the segmentation and protein localization in text files which could be used for further statistical analysis. In addition to computing the quantification, the software also extracts additional features like size of the protein events, the average intensity, gradient and neighborhood information which could be used in higher level analysis.

The software is currently being used as a high throughput analysis tool to quantify protein information in various co-localization studies.

5.1 Level of Detail

A) Parameter setting and Experiment Configuration

The software allows user to enter various experimental parameters to be stored for analysis later on. The software is designed with a greater flexibility in customizing the parameters needed for the algorithms.

B) Modes of operation

The software provides two major modes of operations

1. Foreground mode or interactive mode
2. Batch Mode

In the foreground mode BioQuant can process one image at a time and a visual output is generated for the image that is processed. The main idea of using the software in this

mode is to fine tune the algorithmic parameters for a particular data that is being processed. In batch mode the software could be made to process one directory at a time, usually images in one directory corresponds to data collected under a same experimental group e.g. images collected under same treatment group. Batch mode is generally used to process a large volume of data.

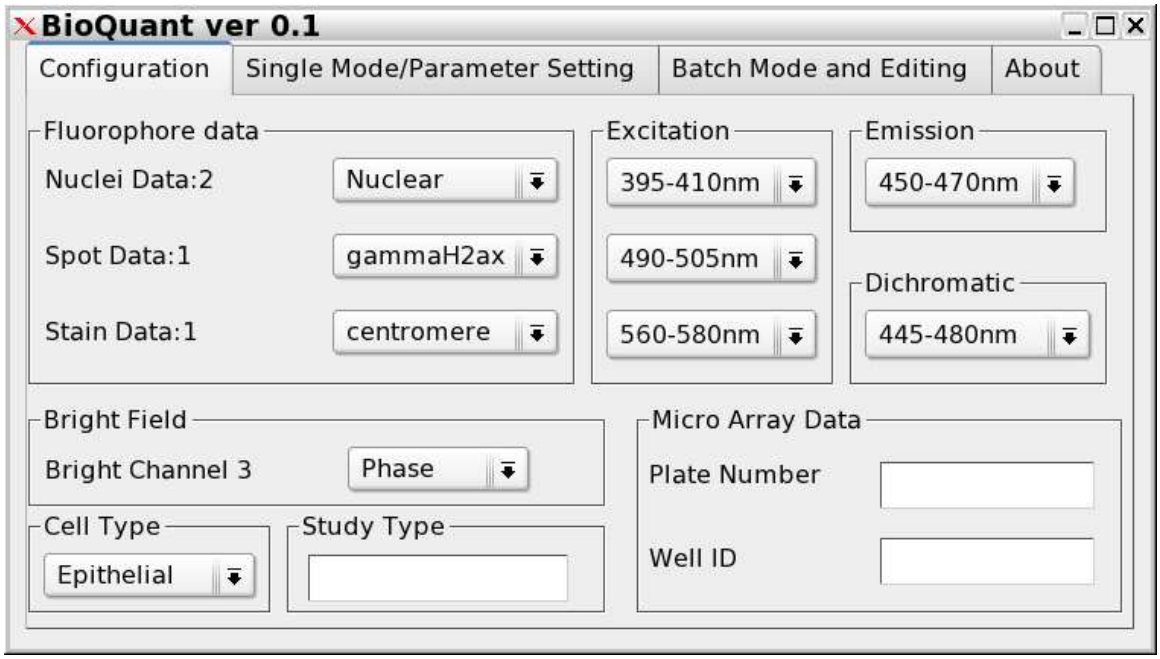


Fig 9. BioQuant: Parameter Configuration panel.

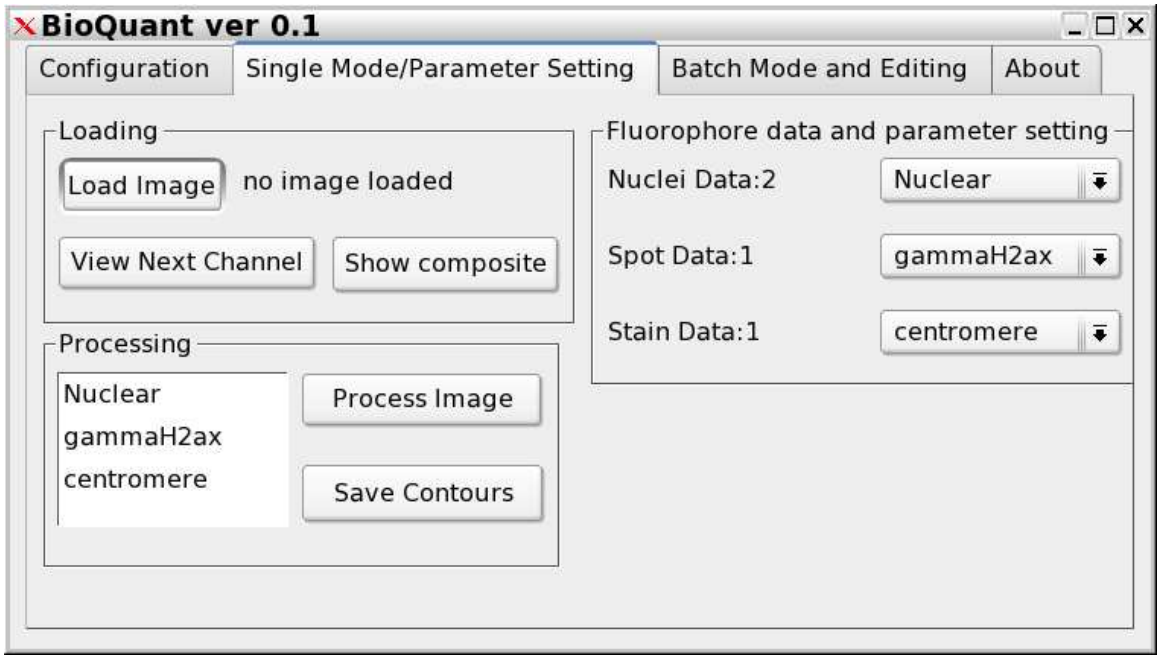


Fig 10: Bioquant: Single mode/ parameter setting panel.

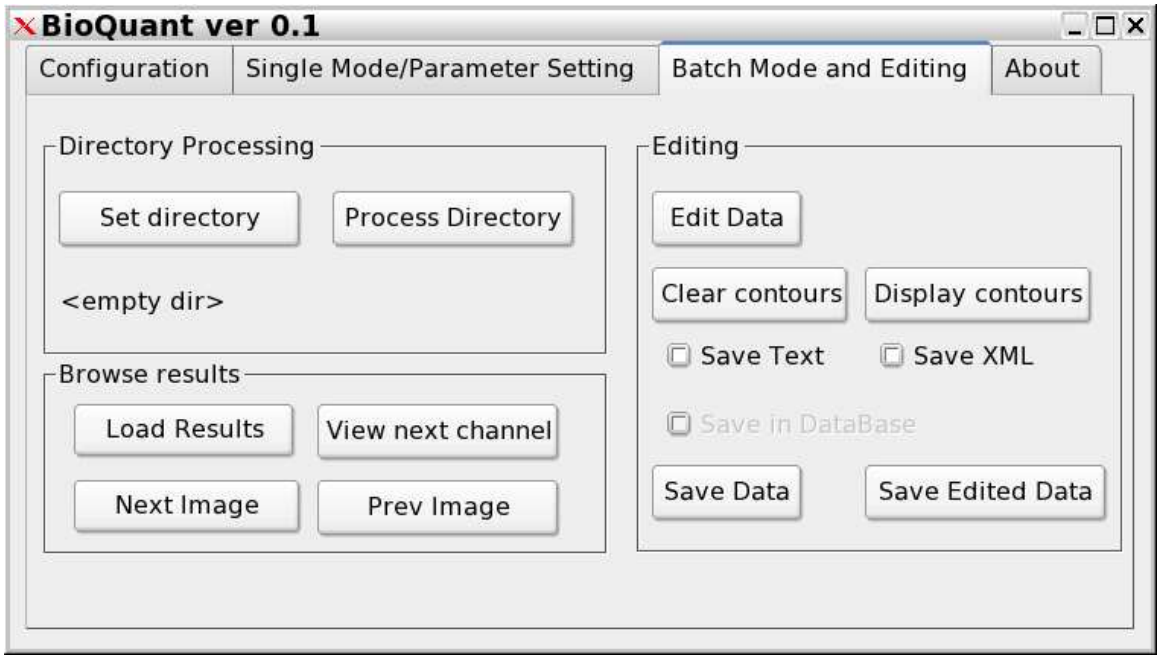
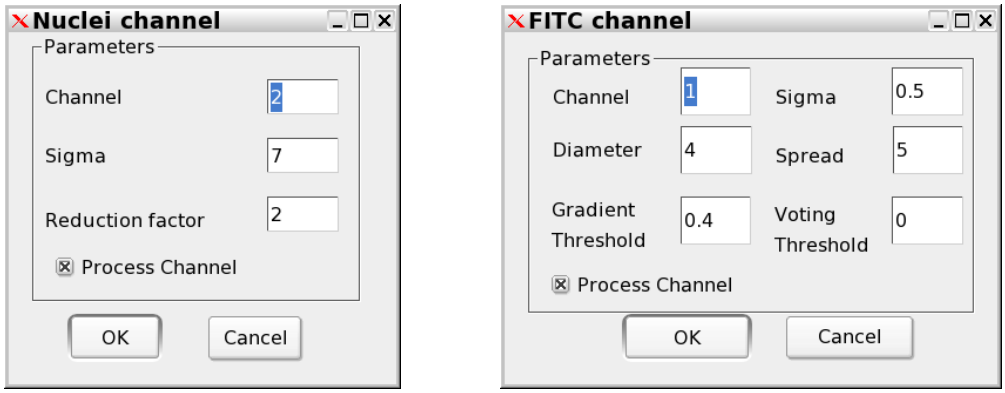


Fig 11. BioQuant: Batch Mode Panel



(a) (b)

Figure 12. BioQuant Parameter: (a)-(b) Algorithmic Parameter Panels

Chapter 6: Experimental Results and Analysis

The system developed here was used in two different studies.

1) Experiment to quantify centrosomal abnormality in breast cancer cells
2) Experiment to quantify and measure protein events in DNA double strand breaks in human epithelial cells subjected to radiation.

6.1 Experiment to study centrosomal abnormality in breast cancer cells

An experiment was conducted to study centrosomal abnormality in breast cancer cells. CA is a rare event which occurs in 2% of normal tissue and 80% of breast cancer cell. Therefore a large sample of data was used to characterize the CA. The cells were treated with different treatment groups and were imaged using fluorescence microscopy. A total of 196 images were processed. The results from the software system, BioQuant were then compared with manual quantification of centrosomal abnormality.

The segmentation and protein localization results are shown in Figures 13 to 19. For each image the segmentation result is displayed in green and protein localization is cyan dots. It should be noted that in certain cases e.g., in Figure 17; compartments 10 and 11 has no decay in intensities between the two touching nuclear regions, well known approaches like watershed method could fail in such cases.

The accuracy of the system in localizing protein events are shown through Figure 13. The centrosomal abnormalities are correctly characterized in cells 14 and 842 and are assigned to appropriate nuclear regions due to correct segmentation of the sub-cellular compartments.

A comparison between manual quantification and automated quantification is shown in Figure 18 and Figure 19. Correlation was computed between the manual and automated quantification and between two investigators. The two facets compared were assigning centrosomes to nuclear regions and counting abnormal centrosomes in a nuclei. From figures 20 and 21 it can be inferred that the manual quantification is subjective to the investigators and the correlation between investigator A and investigator B in both assigning cells to nuclear regions and counting abnormal centrosomes is not ideal. Comparison between manual (average values from two investigators) and automated quantification shows a tighter correlation and the analysis by using an automated approach is no longer subjective but is objective as the algorithms are based on inherent geometries of the measurements being made. Also, it can be noted that the error rate is very marginal and is acceptable measure for high throughput analysis.

Figure 22 shows an example of a failure by using watershed approach for image segmentation in portion of image where the nuclear regions are adjacent to each other. A portion of the image is taken from the image presented in figures 15 and 17, where there are touching nuclear compartments. Watershed approach is intensity based spatial clustering approach and does not use inherent geometry of sub cellular compartments for delineation. It often leads to fragmentation and is sensitive to parameter selection. However by using the geometric based iterative decomposition approach this problem is overcome and the correct delineation of these sub-cellular compartments is shown in figure 15 represented by nuclei id 59, nuclei id 61; and figure 17 represented by nuclei id 10, nuclei id 11.

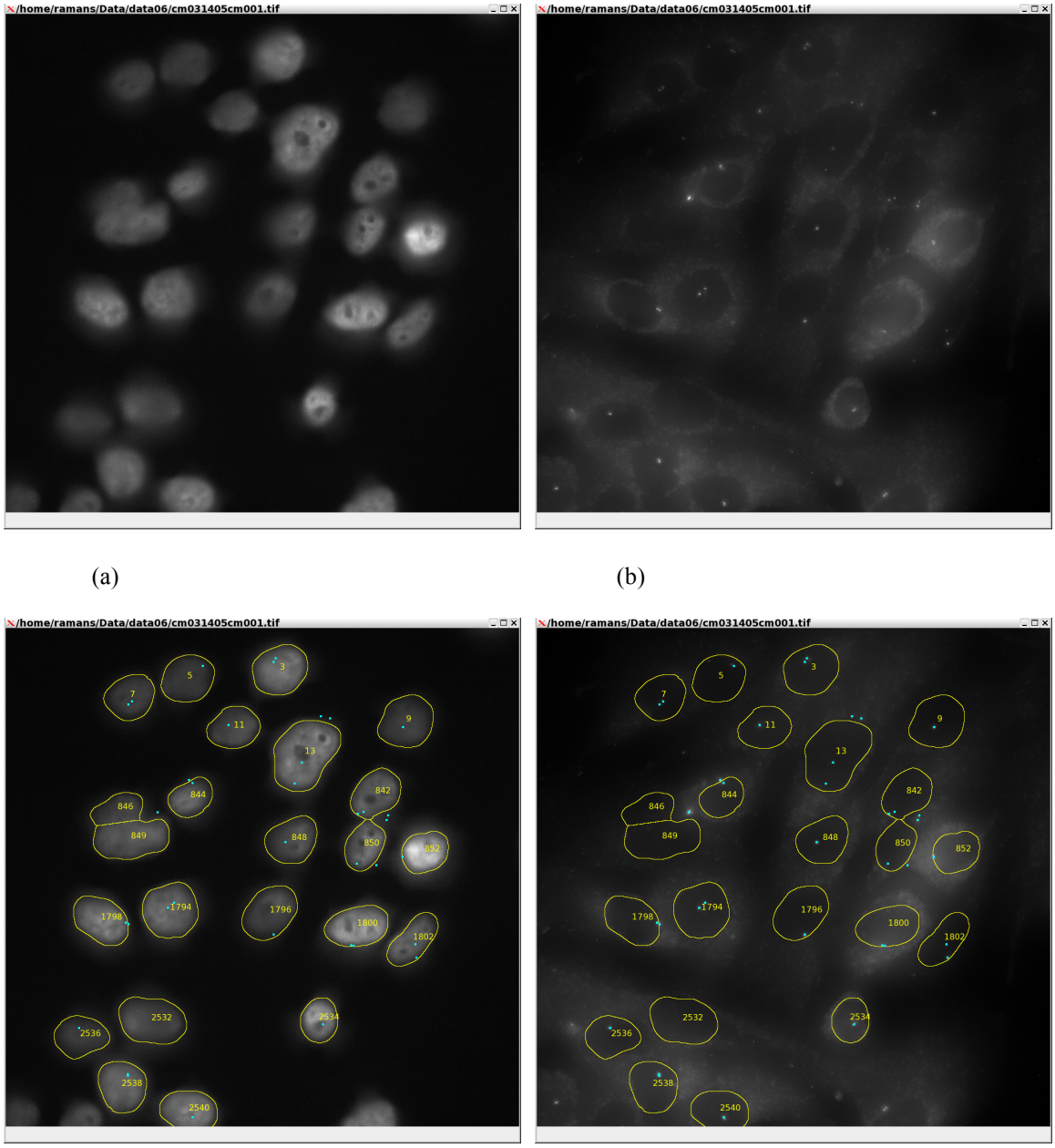


Fig 13. Segmentation and Protein Localization result: (a) Original Image Cellular morphology (b) Protein Events (c)-(d) Result of segmentation and protein localization

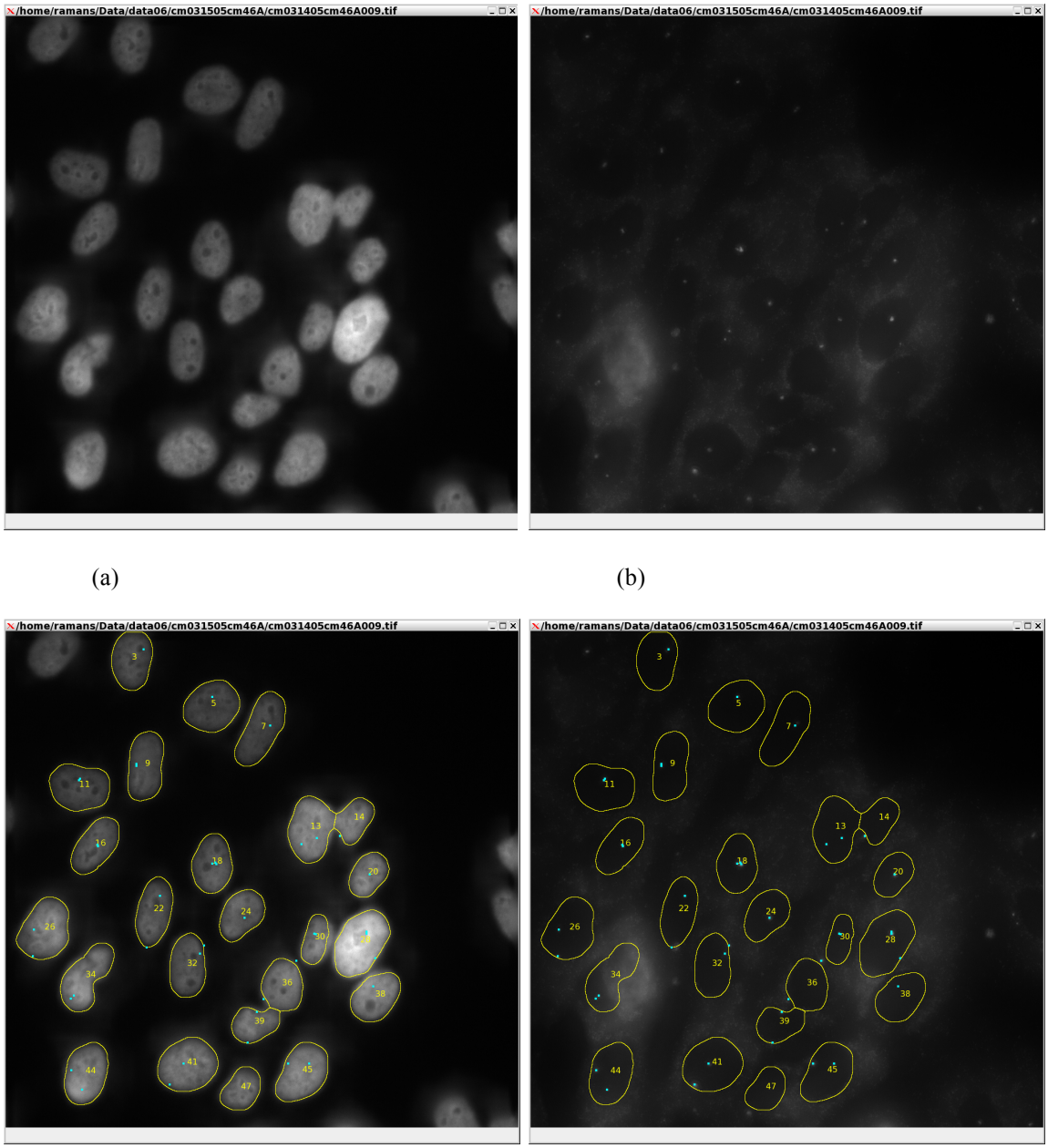


Fig 14. Segmentation and Protein Localization result: (a) Original Image Cellular morphology (b) Protein Events (c)-(d) Result of segmentation and protein localization

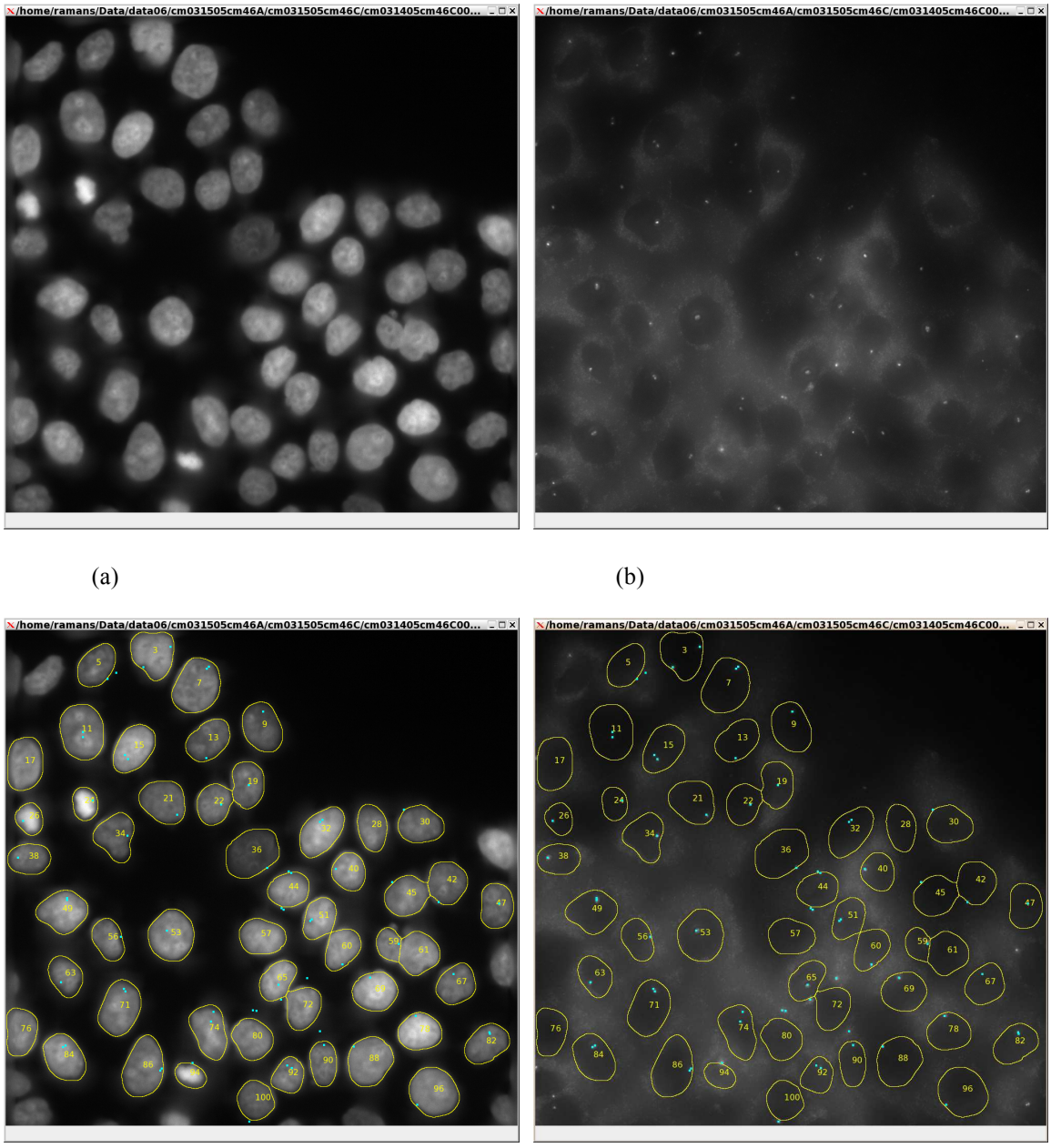


Fig 15. Segmentation and Protein Localization result: (a) Original Image Cellular morphology (b) Protein Events (c)-(d) Result of segmentation and protein localization

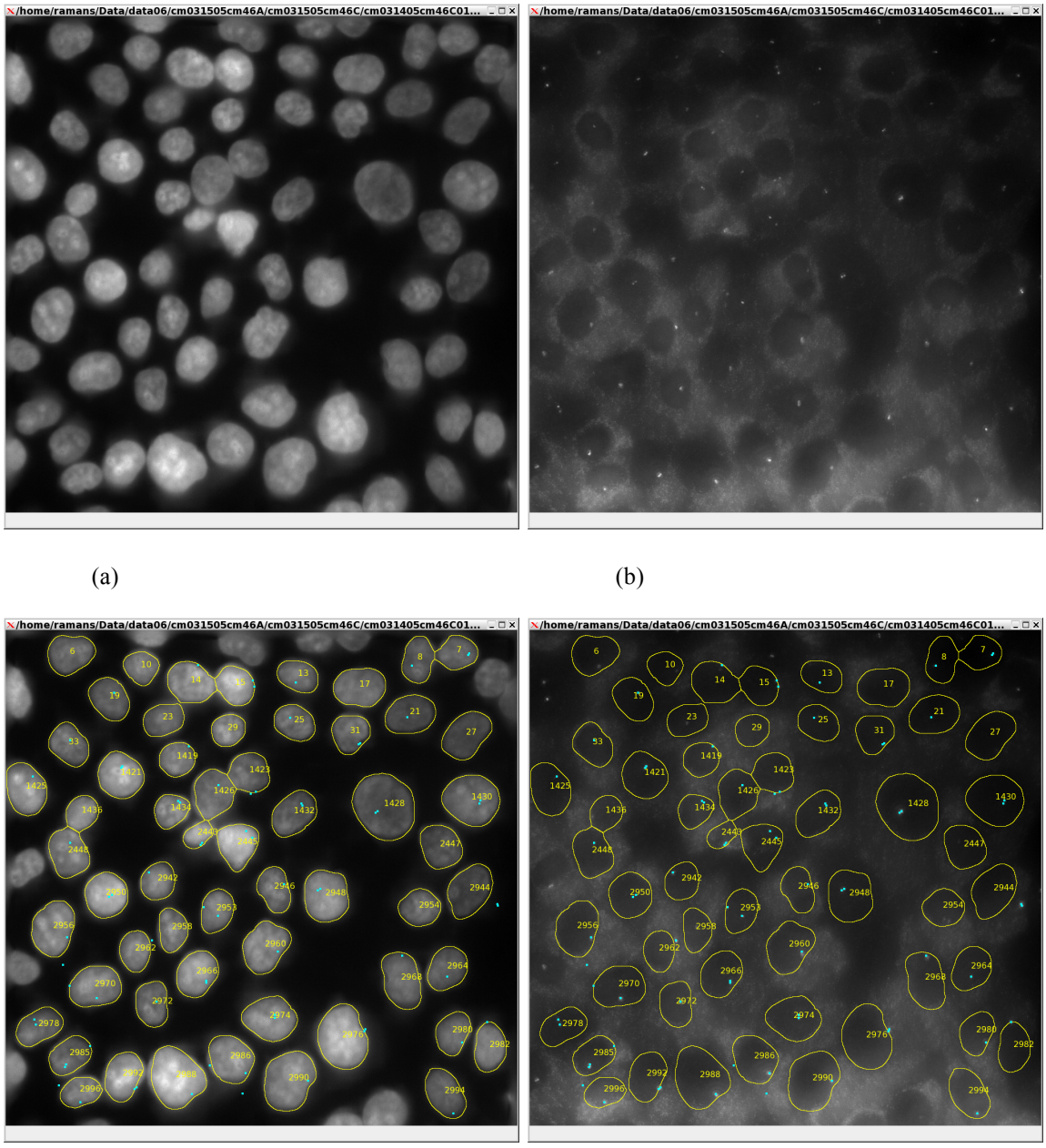
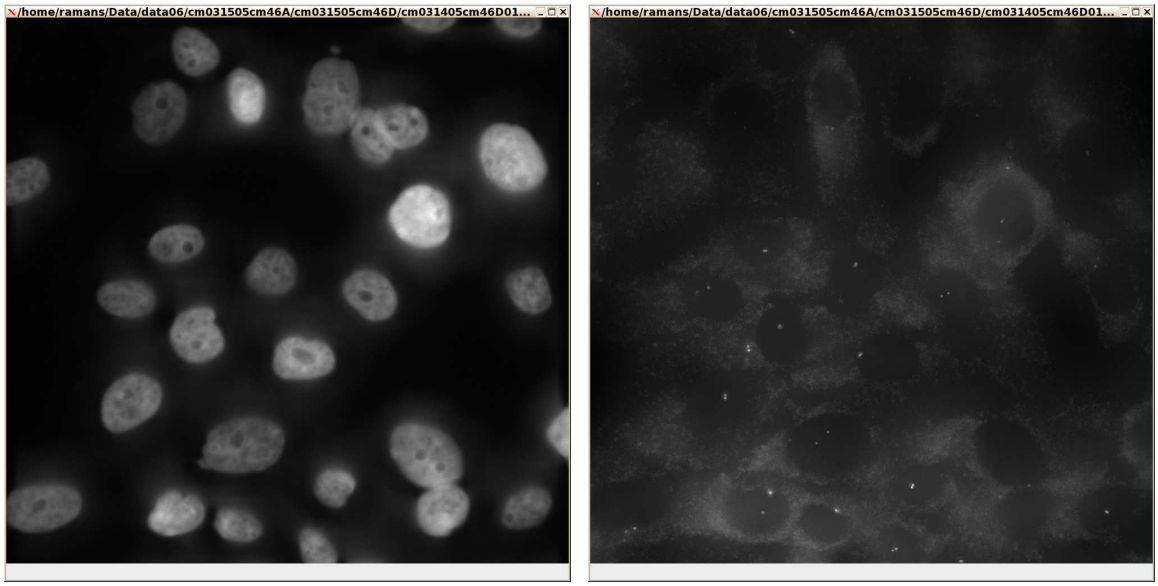
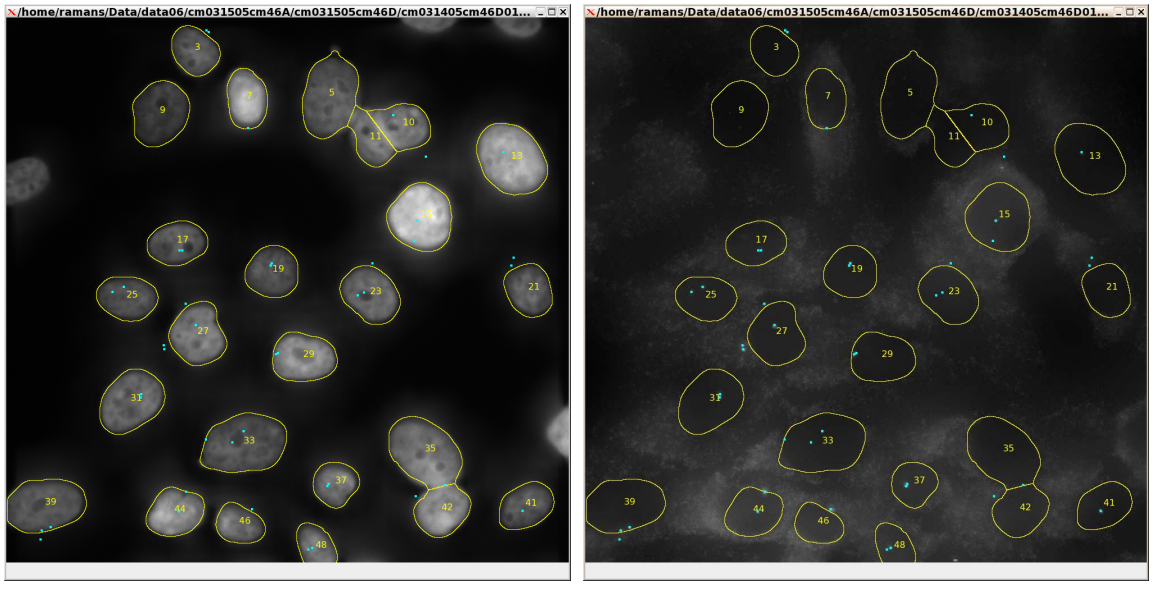


Fig 16. Segmentation and Protein Localization result: (a) Original Image Cellular morphology (b) Protein Events (c)-(d) Result of segmentation and protein localization



(a) (b)



(c) (d)

Fig 17. Segmentation and Protein Localization result: (a) Original Image Cellular morphology (b) Protein Events (c)-(d) Result of segmentation and protein localization

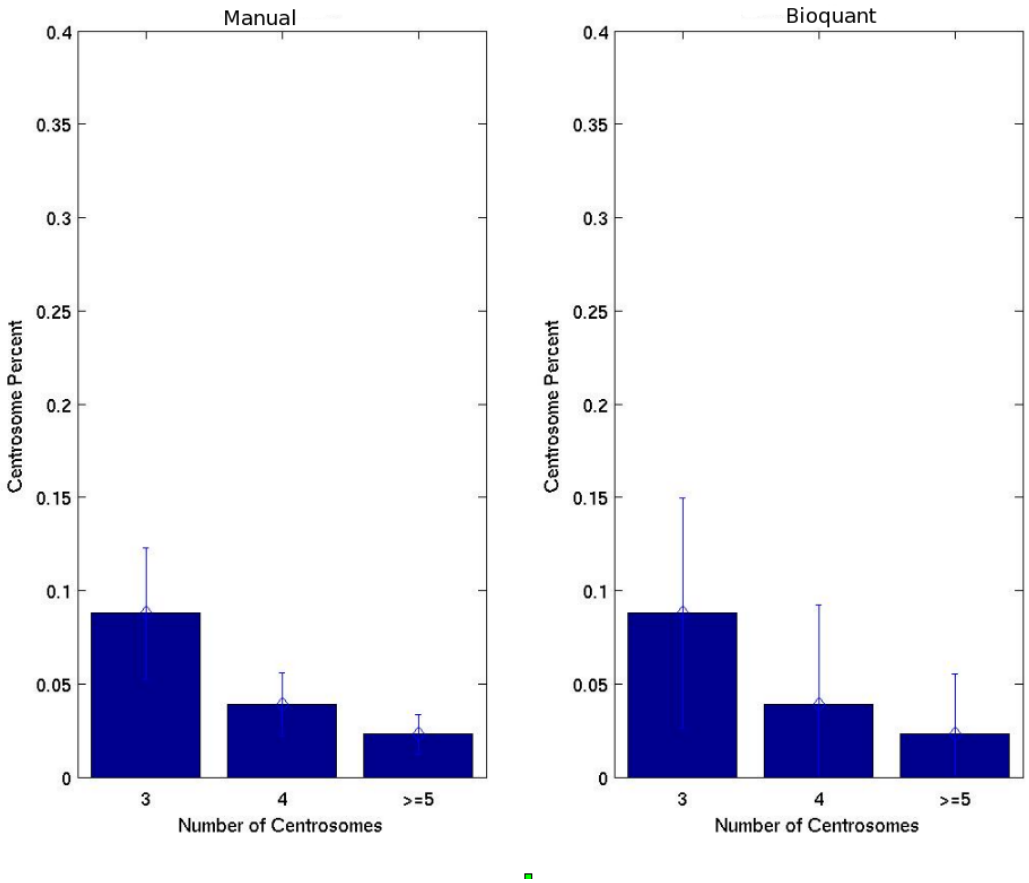


Fig 18. Results Comparison: Manual quantification Vs automated quantification of centrosomal abnormality for treatment group A

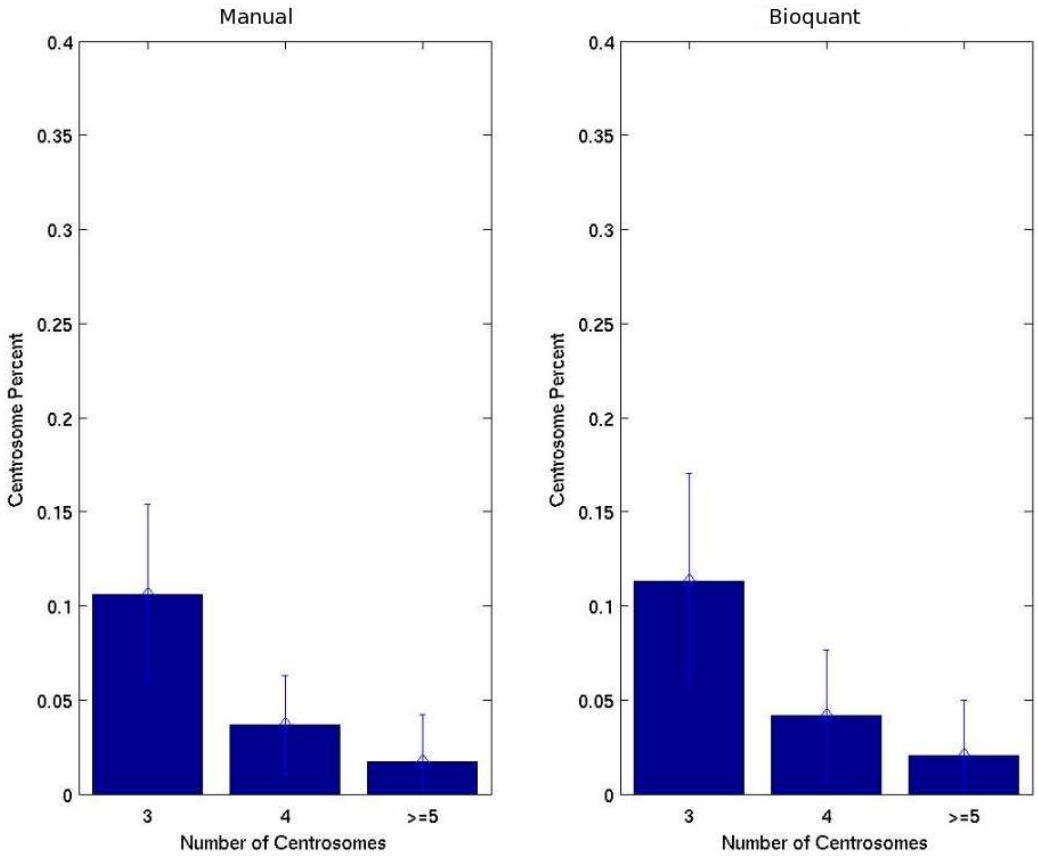
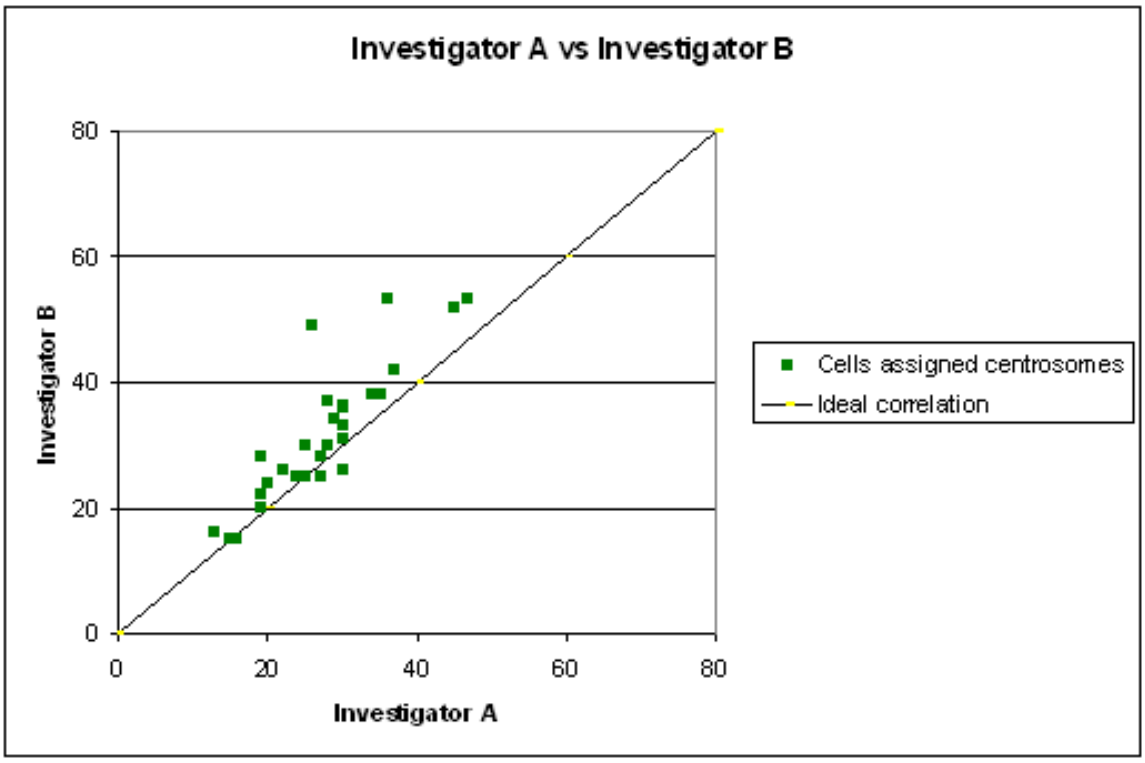
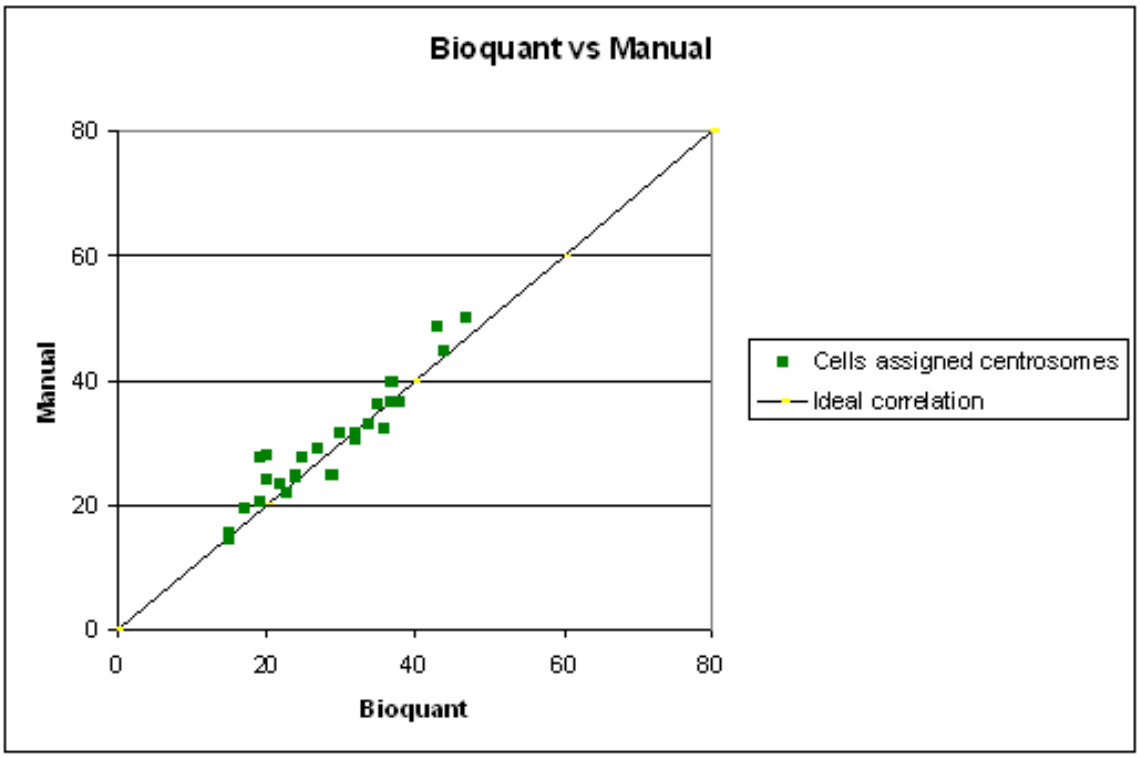


Fig 19 Results Comparison: Manual quantification Vs automated quantification of centrosomal abnormality for treatment group C

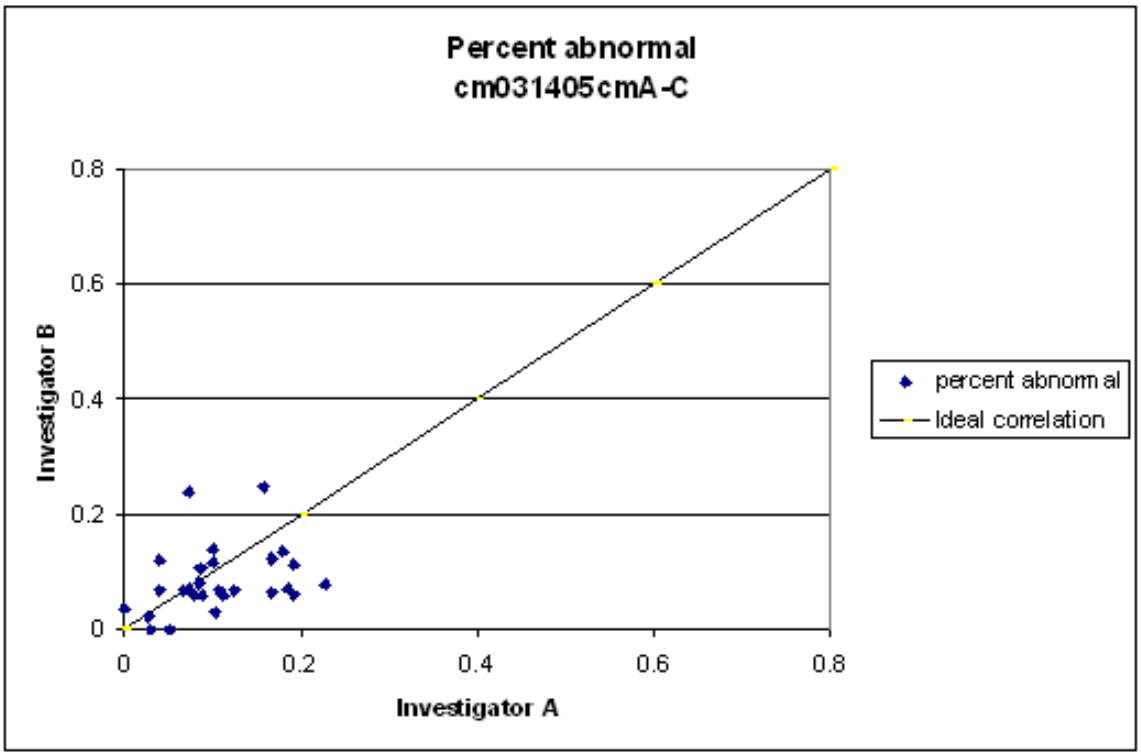

Investigator A vs Investigator B
Investigator B
Investigator A
0
20
40
60
80
Cells assigned centrosomes
Ideal correlation

(a)

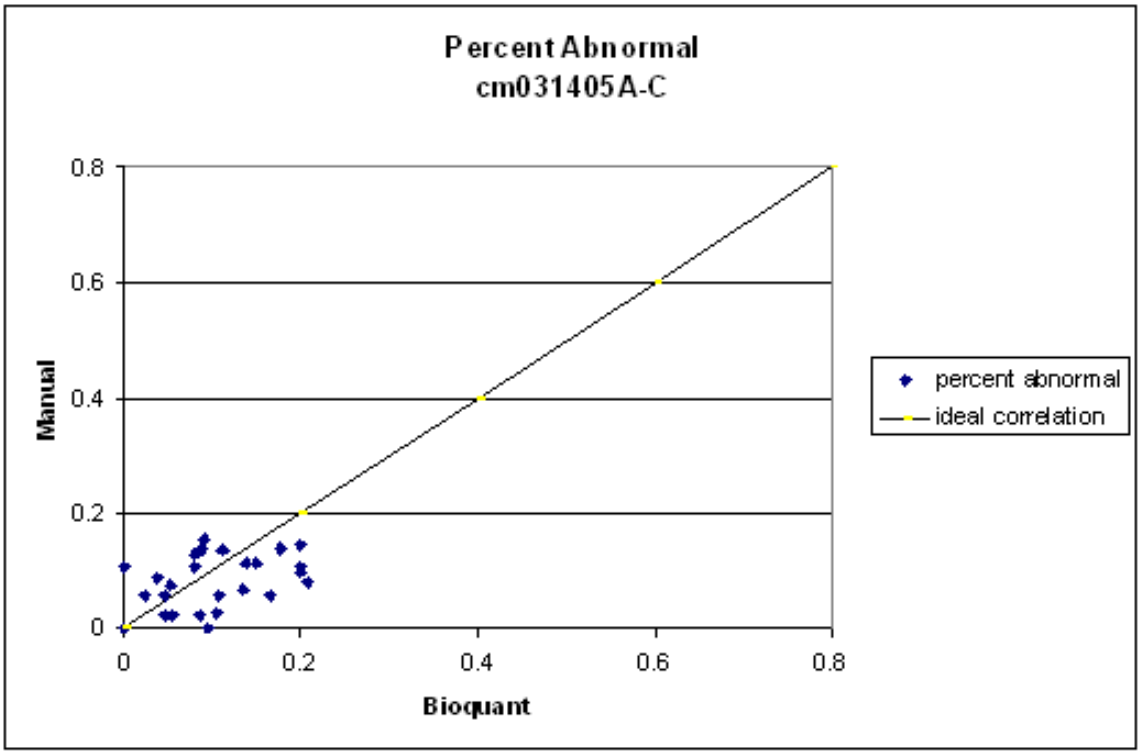


(b)

Fig 20. Results Correlation: Correlation-assigning centrosomes to nuclei (a) Investigator A Vs Investigator B (b) Manual (Average of Investigator A Vs Investigator B).

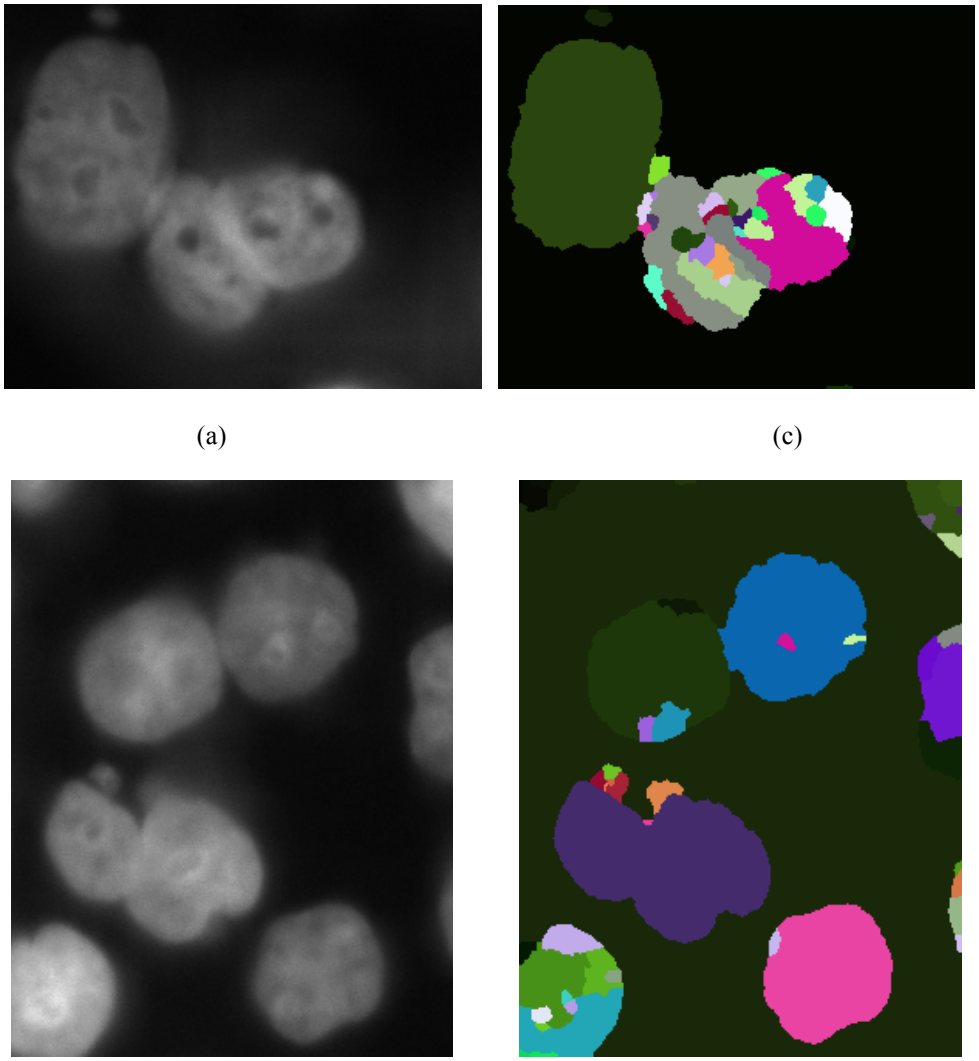

Percent abnormal
cm031405cmA-C
Investigator B
Investigator A
0
0.2
0.4
0.6
0.8
percent abnormal
Ideal correlation

(a)



(b)

Fig 21. Results Correlation: Correlation-detecting abnormal centrosomes (a) Investigator A Vs Investigator B (b) Manual (Average of Investigator A Vs Investigator B)



(a) (c)

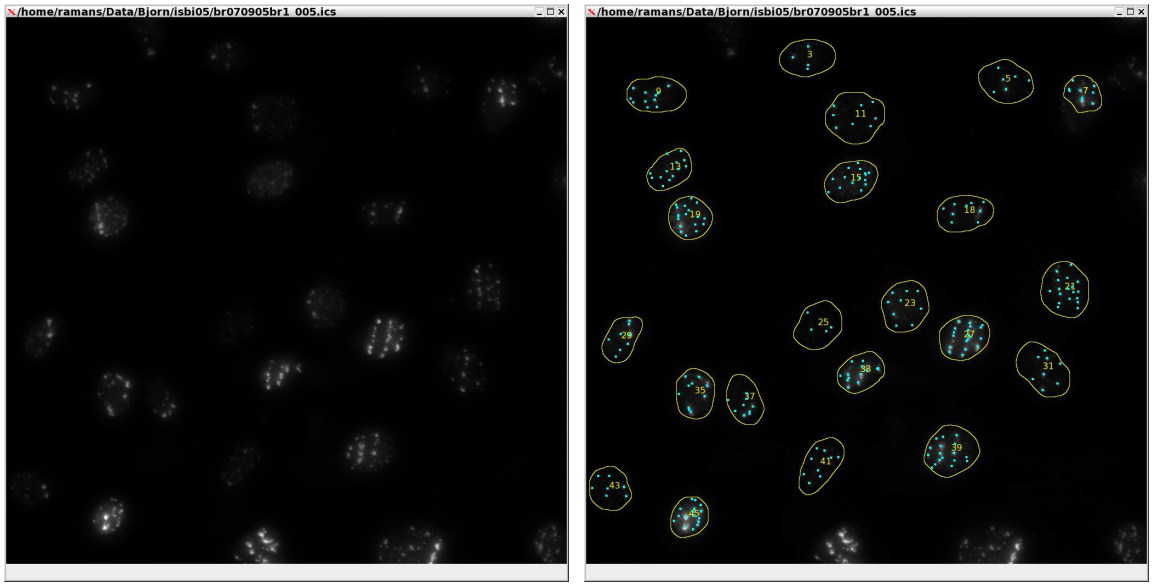
(b) (d)

Fig 22. Watershed Segmentation: (a) – (b) Portion of Original image (c)- (d): Regions segmented by using watershed . Watershed relies on clear differential difference between neighboring objects and as a result in (d), two overlapping nuclei are merged. However, two other nearby objects in the same image are seperated. The parameters selections for these two images were identical and they were chosen to balance the best results in both cases.

6.2 DNA double strand break in human epithelial cells

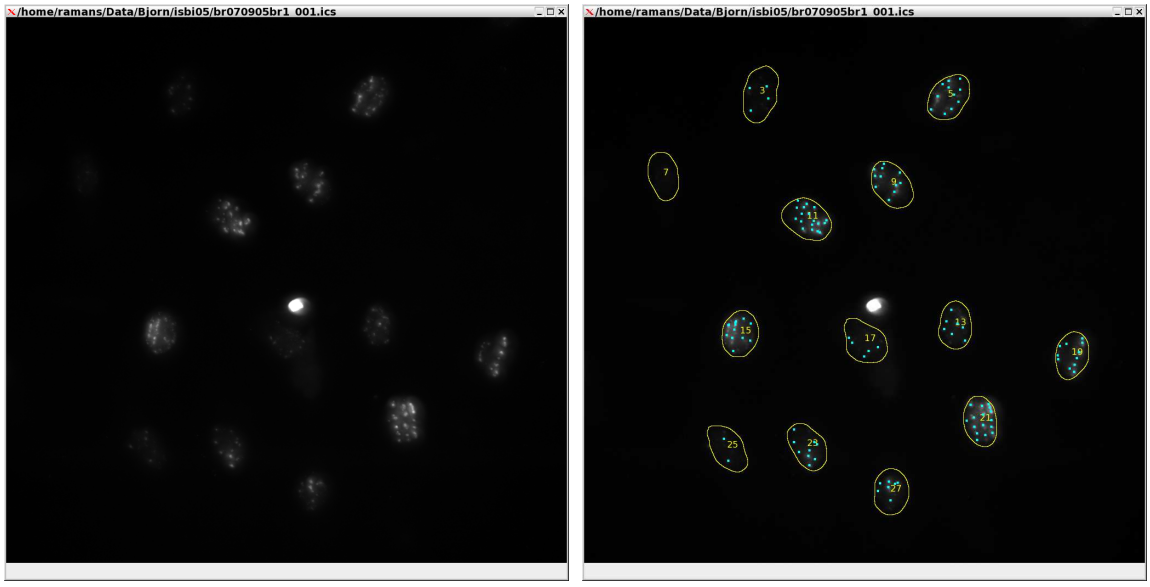
This technique has been applied to foci associated with DNA double-strand breaks (DSBs) in human epithelial cells exposed to high energy (1 GeV/amu) radiations. Such high energy heavy ions (HZE) are present in the galactic cosmic rays (GCR) and is a health concern for space travel. Histone H2AX is phosphorylated in chromatin adjacent to DSBs and can be detected with specific antibodies as punctuated events known as foci. Cell culture slides were irradiated at 0 and 100 centigray of Fe ions at the NSRL facility at Brookhaven National Laboratory, incubated for repair and then stained for phosphorylated H2AX protein. A dataset comprising of 120 images at different time points were analyzed. An example is shown in Figures 23 and 24, and the kinetics of foci formation (repair)--post radiation--over the entire dataset is shown in Figure 25.

Once events are localized, other features such as size and contrast are also computed for other analysis. Of particular interest is pattern of foci formation, which is computed through triangulation, and represented as an attributed graph, shown in Figure 26. This pattern of foci formation is captured using delauany triangulation and could be used to higher level pattern analysis.



(a) (b)

Fig 23. Segmentation and event localization: (a) original image and its corresponding punctate events of γ-H2AX proteins; (b) detected events.



(a) (b)

Fig 24. Segmentation and event localization: (a) original image and its corresponding punctate events of γ-H2AX proteins; (b) detected events.

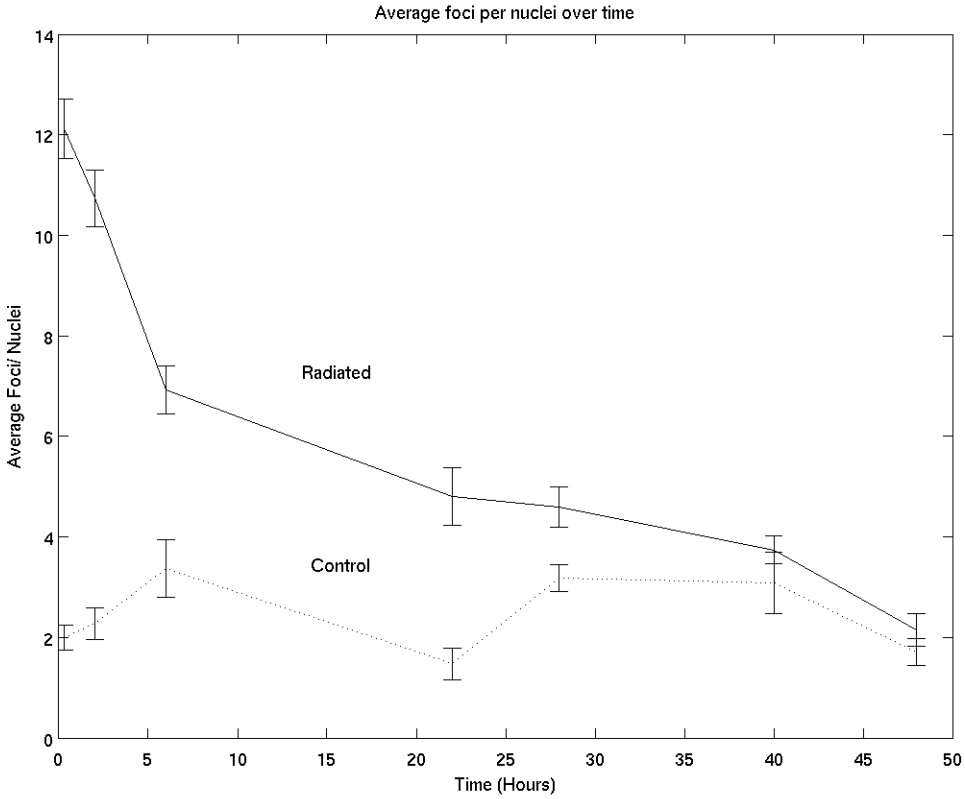
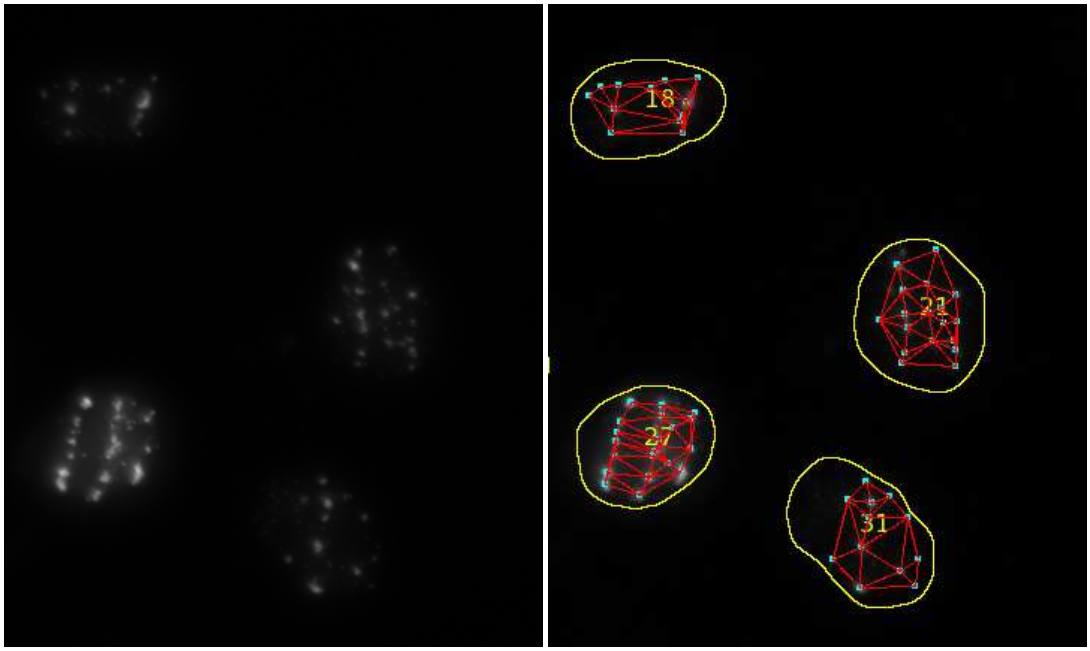


Fig 25. Time Response: Average number of foci (punctate events) per nucleus as a function of time for exposed (solid line) versus control cell lines (dotted lines) indicate dynamics of repair mechanism for exposed cells. Approximately 7-8 images were collected at each time point.



(a) (b)

Fig 26. Foci Pattern: Zoomed portion of image in figure 23 to show the pattern of localized events (a) Original image (b) Pattern of foci represented by delaunay triangulation.

Chapter 7: Conclusion and Future work

The goal of the work presented in this thesis was to accurately segment nuclear compartments and to quantify protein events in cell cultured assays. The work addresses large variations in signatures of the sub-celluar compartments and protein events that occur in cell cultured assays and accurately quantifies the signals inspite of varying degrees of complexities.

Segmentation of nuclear compartments provides the context for measuring the protein events. A novel approach for segmentation by leveraging geometric features of the sub-cellular compartments was presented. This algorithm provided a better segmentation results over well known segmentation approaches. An iterative voting technique is used to localize punctuate protein events which leverages the radial symmetry of punctuate protein events was implemented. It provides robust and accurate result over varying scales and intensities.

The algorithms presented was used to quantify protein events in two studies

1) Quantifying Centrosomal abnormality,
2) Study mechanism of DNA double strand breaks in human epithelial cells

The results discussed showed that the system can make robust and accurate measurement and proved to be a valuable tool for high throughput analysis. These algorithms can able to extend towards a large variety of phenotypic studies.

The performance of the current system can be enhanced further by using dynamic programming in iterative decomposition. Currently, there is a time overhead in

calculating all possible hypotheses for a potential decomposition which has a higher order of magnitude in time complexity as the number of sub-cellular compartments to be delineated increases. This could be reduced by using dynamic programming approach where at each iteration an optimum partition could be extracted and the cost function encompassing the convexity could be propagated to get a globally optimum decomposition.

APPENDIX A

Decomposition Algorithm

1. Localize positive curvature maxima along the contour

2. Set initial number of compartments n:= 2

3. Construct a set of all valid configurations of compartments by connecting valid pairs of positive curvature maxima satisfying the anti parallel, nonintersecting and convexity constraints

4. Evaluate cost of each configuration as defined by the convexity measure defined in Equation 2.

5. Increment the compartment count n:=n+1 and repeat steps 3 and 4 until there is at least one configuration (e.g., a set of partitions where each partition is convex)

6. Select the configuration with the least cost function

APPENDIX B

Iterative Voting

1. *Initialize the parameters*: Initialize r_{min}; r_{max}; Δ_{max}, and a sequence $\Delta_{max} = \Delta_N < \Delta_{N-1} < \ldots < \Delta_0 = 0$. Set n := N, where N is the number of iterations, and let $\Delta_n = \Delta_{max}$. Also fix a low gradient threshold, ε and a kernel variance, σ, depending on the expected scale of salient features.

2. *Initialize the saliency feature image*: Define the feature image F(x; y) to be the local external force at each pixel of the original image. The external force is often set to the gradient magnitude or maximum curvature depending upon the type of saliency grouping and the presence of local feature boundaries.

3. *Initialize the voting direction and magnitude*: Compute the image gradient, and its magnitude, Define a pixel subset S := {(x; y) | magnitude of gradient > ε }. For each grid point (x; y) in S, define the voting direction to be

$$\alpha(x, y) := -\frac{\nabla I(x, y)}{||\nabla I(x, y)||}$$

4. *Compute the votes*: Reset the vote image V (x; y; r_{min}; r_{max}; Δn) = 0 for all points (x; y). For each pixel (x; y) in S, update the vote image as follows:

$$V(x, y; r_{\min}, r_{\max}, \Delta_n) := V(x, y; r_{\min}, r_{\max}, \Delta_n) + \sum_{(u,v)\in A(x,y;r_{\min},r_{\max},\Delta_n)} F(x - \tfrac{w}{2} + u, y - \tfrac{h}{2} + v) K(u, v; \sigma, \alpha, A),$$

where w = max(u) and h = max(v) are the maximum dimensions of the voting area.

5. *Update the voting direction*: For each grid point (x; y) in S, revise the voting direction. Let

$$(u^*, v^*) = \arg \max_{(u,v) \in A(x,y;r_{\min},r_{\max},\Delta_n)} V(u, v; r_{\min}, r_{\max}, \Delta_n)$$

Let $d_x = u^* - x$; $d_y = v^* - x$, and

$$\alpha(x, y) = \frac{(d_x, d_y)}{\sqrt{d_x^2 + d_y^2}}$$

6. *Refine the angular range*: Let n := n - 1, and repeat steps 4-6 until n = 0.

7. *Determine the points of saliency*: Define the centers of mass or completed boundaries by thresholding the vote image:

$$C = \{(x, y) \mid V(x, y; r_{\min}, r_{\max}, \Delta_0) > \Gamma_v\}$$

REFERENCES

1. R. Malladi and J. Sethian. A unified approach to noise removal, image enhancement, and shape recovery. *IEEE Transactions on Image Processing*, 5(11):1554-1568, 1995.

2. G. Medioni, M.S. Lee, and C.K. Tang. A Computational Framework for Segmentation and Grouping. *Elsevier*, 2000.

3. R. Murphy. Automated interpretation of subcellular locatoin patterns. *In IEEE Int. Symp. on Biomedical Imaging*, pages 53 - 56, April 2004.

4. C. et. al. Ortiz De Solorzano. Segmentation of nuclei and cells using membrane protein. *Journal of Microscopy*, 201:404 - 415, March 2001.

5. B. Parvin, Q. Yang, G. Fontenay, and M. Barcellos-Hoff. Biosig: An imaging bioinformatics system for phenotypic analysis. *IEEE Transactions on Systems, Man and Cybernetics*, 33(B5):814 - 824, October 2003.

6. D. Reisfeld, H. Wolfson, and Y. Yeshurun. Context-free attentional operators: The generalized symmetry transform. *IJCV*, 14(2):119 -130, March 1995.

7. D. Reisfeld and Y. Yeshurun. Preprocessing of face images: Detection of features and pose normalization. *CVIU*, 71(3):413- 430, September 1998.

8. J. L. Salisbury. The contribution of epigenetic changes to abnormal centrosomes and genomic instability in breast cancer. *Journal of Mammary Gland Biology and Neoplasia*, 6(2):203-12, April 2001.

9. J. L. Salisbury, A. B. D'Assoro, and W. L. Lingle. Centrosome ampli_cation and the origin of chromosomal instability in breast cancer. *Journal of Mammary Gland Biology and Neoplasia*, 9(3):275-83, July 2004.

10. G. Sela and M.D. Levine. Real-time attention for robotic vision. *Real-Time Imaging*, 3(3):173-194, June 1997.

11. L. Vincent and P. Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *PAMI*, 13(6):583-598, June 1991.

12. Q. Yang and B. Parvin. Harmonic cut and regularized centroid transform for localization of subceullar structures. *IEEE Transactions on Biomedical Engineeing*, 50(4):469-475, April 2003.

13. Q. Yang and B. Parvin. Perceptual organization of radial symmetries. *In Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 320-325, 2004.